DOCUMENT RESUME

ED 356 386 CE 063 477

AUTHOR Wadleigh, Sandra L.; And Others
TITLE Testing and the Curriculum: Proceed with Caution.
PUB DATE 93
NOTE 11p.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Allied Health Occupations Education; *Educational Testing; High Schools; Nontraditional Education; Nurses Aides; Postsecondary Education; *Reading Tests; *Student Placement; *Testing Problems; Test Reliability; Test Theory; *Test Use; Test Validity
IDENTIFIERS *Assessment of Student Skills for Entry Transfer; *Nelson Denny Reading Tests

ABSTRACT

A study compared the performance of 44 applicants seeking admission to an alternative high school (n=19) and nursing assistant program (n=25) at a Wisconsin postsecondary institution on the Assessment of Student Skills for Entry Transfer (ASSET) test and the Nelson-Denny Reading Test. (Applicants who did not achieve a minimum score on ASSET then took the Nelson-Denny.) The hypothesis was that there would be no significant differences in applicants' performance as shown by the raw scores on the two tests. Three test cycles were scheduled. All 44 subjects took ASSET in the first test cycle. Subjects who were determined by a coin toss to be tested in cycle two with the Nelson-Denny and cycle three with ASSET were labeled Group 1 (n=21). Those who were determined by coin toss to be tested in cycle two with ASSET and cycle three with the Nelson-Denny were labeled Group 2 (n=23). A dependent t-test was used to analyze data. Students' performance on ASSET was significantly different from their performance on the Nelson-Denny. Students' performance on ASSET was also significantly different from their performance on a retest with the ASSET. The hypothesis was rejected in four of the five comparisons between performance on the ASSET and performance on the Nelson-Denny. It was recommended that multiple tests used for educational decisions such as placement must measure the same things and must measure what the curriculum deals with. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED356386

TESTING AND THE CURRICULUM: PROCEED WITH CAUTION

Sandra L. Wadleigh

Fox Valley Technical College

Jupian J. Leung

University of Wisconsin-Oshkosh

Robert A. McKeag

University of Wisconsin-Oshkosh

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JJ Leung

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CE063477

BEST COPY AVAILABLE

Testing and the Curriculum: Proceed With Caution

In the spring, 1992, issue of Forward, the WASCD Journal, Karl Hertz and Howard Stone raised some very important questions regarding the role of assessment in the development (or retardation) of curriculum.

Of course, the questions raised by Karl Hertz and Howard Stone, that is, is assessment the means to an end or the end in itself? and what shape should assessment take? are not new questions. They have been raised before. Yet they seem to take on a fresh look, and a new urgency today. In Wisconsin, we have the "Third Grade Test" in reading in place while other state-wide assessments are coming in the very near future. Today, we do seem to be viewing the assessment route as a useful route to follow in a school improvement or curriculum development plan.

We agree that assessment can be a useful tool to improve school performance and validate what is working. We would, however, like to suggest some caution in using the tool based on a study we recently completed which involved the use of assessment procedures to determine placement of applicants into a curriculum.

Our study grew out of some observed inconsistencies noted over a three-year period in the placement of applicants into a post-secondary nursing assistant and alternative high school program. Applicants were tested in reading using the ASSET test developed by the American College Testing program. If a minimum performance was not achieved on this test, applicants were re-tested using a different reading test, the Nelson-Denny Reading

Test developed by the Riverside Publishing Company in Chicago. This should be a good procedure: it supplies more data and gives applicants more of a chance to succeed. It is a good procedure if the two tests are equivalent, that is, if they are both measuring the same thing. If they are not really equivalent, that is, if they are not measuring reading skills, or have different views of what constitutes reading skills (the point raised by Howard Stone in the spring, 1992 issue of Forward), then the procedure is flawed, and may not even be fair to applicants, schools, or the curriculum.

We decided to check out the two tests in question to see if they were equivalent by comparing the performance of 44 applicants seeking admission to the alternate high school and nursing assistant programs at a post-secondary institution in Wisconsin. Of these 44 applicants, 19 were alternate high school applicants and 25 were nursing assistant applicants.

Hypothesis

In this study, it was hypothesized that there would be no significant differences in applicants' performance as shown by the raw scores on the two tests under study. Raw scores were used in this study because that was the practice then in place at the institution in which the study was conducted. This hypothesis was developed from the expectation that if the two reading tests under investigation are equivalent, that is, if they are measuring the same skills, then applicants' performance should be the same on both tests. The expectation was also that if the two reading tests were not equivalent, that is, if they

were not measuring the same skills, then applicants' performance would be different.

Procedure

In part because of the institutional admission procedures, three test cycles were scheduled for testing the subjects in the study. All 44 subjects took the ASSET in the first test cycle. The test administered in the second test cycle, ASSET or Nelson-Denny, was determined randomly for each applicant by a flip of the coin toss. The test administered to the subjects in cycle three was the one not administered to them in cycle two. Subjects who were determined by the coin toss to be tested in cycle two with the Nelson-Denny and cycle three with the ASSET were labelled Group One. Those who were determined by the coin toss to be tested in cycle two with the ASSET and cycle three with the Nelson-Denny were labelled Group Two. Group One consisted of 21 and Group Two consisted of 23 subjects. Each group consisted of both alternate high school and nursing assistant applicants. The Group One and Group Two subjects were tested separately. All testing was done in an enclosed room with subjects seated at individual tables. To avoid the possible loss of subjects and other complications, a time period of two weeks was arbitrarily chosen to separate the first test cycle and the subsequent two test cycles. The first author administered the tests to all subjects in all test cycles. The test administered in the second test cycle was either the ASSET or the Nelson-Denny, depending on the group to which the subjects belonged, as described previously. During the third test cycle, which

immediately followed the second, subjects were administered the test not previously taken during the second cycle, as previously noted. This procedure is illustrated below.

Test Procedures

Cycle	Test	Description
One	ASSET	All applicants, nursing assistant as well as alternative high school
Two	ASSET or Nelson-Denny	Determined by coin flip as to which test applicants would take
Three	ASSET or Nelson-Denny	Applicants took the test they had not taken in cycle two

The first author read the printed directions for each test to each group of subjects. Each test cycle was timed and the test used in each test cycle was administered in a twenty-minute time frame, which was the suggested time limit for both tests. The two groups of applicants were tested one day apart in the same week during the same time of day. All tests were hand scored by the first author for each subject in each group.

Data Analysis

As noted previously, the raw scores from the tests were used for data analysis in conformance with the practice then in place at the institution for placing applicants into the curriculum. A dependent *t*-test was used to analyze the data.

The first comparison we made was on the raw scores of the 21 nursing assistant and alternate high school applicants of Group One who were tested in cycle one with the ASSET and cycle two with the Nelson-Denny. The dependent *t*-test showed $t(20) = 3.62$, $p < .01$. This result shows that the applicants' performance was

significantly higher on the Asset (mean=21.76, standard deviation=5.25) than on the Nelson-Denny (mean=18.09, standard deviation=5.28).

The second comparison we made was on the raw scores of the 21 nursing assistant and alternate high school applicants of Group One who were tested in cycle one with the ASSET and cycle three with the ASSET. The dependent *t*-test showed $t(20) = -.58$, $p > .05$. This result shows that the applicants' performance did not differ significantly between the ASSET in cycle one (mean=21.76, standard deviation=5.25) and the ASSET in cycle three (mean=22.38, standard deviation=6.32).

The third comparison was made on the raw scores of the 23 nursing assistant and alternate high school applicants of Group Two who were tested in cycle one with the ASSET and cycle three with the Nelson-Denny. The dependent *t*-test showed $t(22) = 1.42$, $p > .05$. This result shows that the applicants' performance did not differ significantly between the ASSET in cycle one (mean=19.65, standard deviation=5.67) and the Nelson-Denny in cycle three (mean=17.91, standard deviation=6.75).

The fourth comparison was made on the raw scores of the 23 nursing assistant and alternate high school applicants of Group Two who were tested in cycle one with the ASSET and cycle two again with the ASSET. The dependent *t*-test showed $t(22) = -2.35$, $p < .05$. This result shows that the applicants' performance was significantly lower on the ASSET in cycle one (mean=19.65, standard deviation=5.67) than on the ASSET in cycle two (mean=21.69, standard deviation=7.24).

The fifth comparison we made was on the raw scores of the 19 alternate high school applicants from both Group One and Group Two who were tested with the ASSET in cycle one and the Nelson-Denny in either cycle two or cycle three. The dependent *t*-test showed $t(18) = 2.63$, $p < .05$. This result shows that the applicants' performance was significantly higher on the ASSET (mean=18.00, standard deviation=5.29) than on the Nelson-Denny (mean=14.52, standard deviation=6.51).

The sixth comparison we made was on the raw scores of the 25 nursing assistant applicants from both Group One and Group Two who were tested with the ASSET in cycle one and the Nelson-Denny in either cycle two or cycle three. The dependent *t*-test showed $t(24) = 2.02$, $p < .05$. This result shows that the applicants' performance was significantly higher on the ASSET (mean=22.68, standard deviation=4.86) than on the Nelson-Denny (mean=20.64, standard deviation=4.05).

The seventh comparison was made on the raw scores of all 44 nursing assistant and alternate high school applicants from both Group One and Group Two who were tested with the ASSET in cycle one and the Nelson-Denny in either cycle two or cycle three. The dependent *t*-test showed $t(43) = 3.62$, $p < .01$. This result shows that the applicants' performance was significantly higher on the ASSET (mean=20.65, standard deviation=5.51) than on the Nelson-Denny (mean=17.54, standard deviation=6.31).

The eighth comparison was made on the raw scores of all 44 nursing assistant and alternate high school applicants from both Group One and Group Two who were tested with the ASSET in cycle

one and again in cycle two or cycle three. The dependent *t*-test showed $t(43) = -2.95$, $p < .01$. This result shows that the applicants' performance was significantly lower on the ASSET in cycle one (mean=20.65, standard deviation=5.51) than on the ASSET in either cycle two or cycle three (mean=22.77, standard deviation=7.08).

The findings described above thus indicate that, by and large, students' performance on the ASSET is significantly different from their performance on the Nelson-Denny. Interestingly, even students' performance on the ASSET is significantly different from their performance on a re-test with the ASSET, as shown in the fourth and eighth comparisons noted above.

Discussion and Suggestions

The hypothesis in our study was rejected in four of the five comparisons between performance on the ASSET and performance on the Nelson-Denny test (comparisons one, three, six, seven noted above). It's not clear at all that these two reading tests are equivalent for purposes of placing applicants into a post-secondary nursing assistant or alternate high school curriculum. These results raise several questions and certainly underscore the points posed by Hertz and Stone in the spring, 1992 issue of Forward.

If assessment and testing is going to be used to make educational decisions (and we feel certain that this will continue to be the case) then it is very important to take a hard look at the assessment procedure and instruments being used. It

seems quite clear to us that tests, even in the same curricular area, do not always measure the same things. We need to take a hard look at what each of those tests is measuring and we need to ask ourselves if what the test measures is what the curriculum is dealing with, and if what the test measures is what children are being asked to learn.

We probably also need to take a hard look at our assessment procedures to see if they address what we might call "curriculum variables." For example, a reading test that measures reading skills may be very inappropriate to use in a reading program in which "process" or "whole language" is emphasized.

When we make decisions about students based on test results we also need to proceed with caution. In moving students to a different level of curriculum or in admitting them into a certain curriculum we can almost guarantee failure if the placement decision is faulty. And our placement decision might be faulty if our assessment is flawed.

In considering state-wide assessment plans, we probably also need to be cautious in light of our findings. We need to be sure that assessment starts by determining the needs of students first. Sometimes the test we plan to use dictates our curriculum and colors our decision making. In this case the assessment procedure erodes our freedom to develop curriculum to meet the needs of students---it is the needs of assessment that we are thinking of instead. This strikes us as a "tail wagging the dog" situation.

Assessment is an important and useful tool in education, but it is a tool that needs to be carefully considered and examined. As Stone and Hertz noted in their comments, we need to think about the kind of assessment that will do the job we wish to have done and we need to ask ourselves whether assessment is a means to that end or an end in itself.

References

Hertz, Karl V. "Curriculum and Assessment--The Means or the End." Forward, 16, Spring, 1992.

Stone, Howard L. "Alternative Assessments: The Missing Link!" Forward, 16, Spring, 1992.